RSAC 2025 Conference

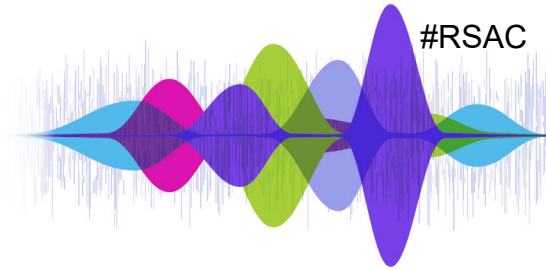
Many Voices.
One Community.

SESSION ID: SAT-M06

The Machines Are Learning, But Are We?

Steve Wilson

Chief Product Officer
Exabeam



Disclaimer

Presentations are intended for educational purposes only and do not replace independent professional judgment. Statements of fact and opinions expressed are those of the presenters individually and, unless expressly stated to the contrary, are not the opinion or position of RSA Conference LLC or any other co-sponsors. RSA Conference LLC does not endorse or approve, and assumes no responsibility for, the content, accuracy or completeness of the information presented.

Attendees should note that sessions may be audio- or video-recorded and may be published in various media, including print, audio and video formats without further notice. The presentation template and any media capture are subject to copyright protection.

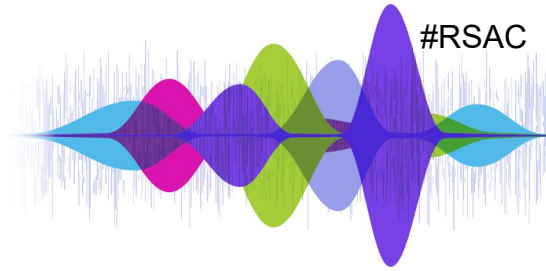
© 2025 RSA Conference LLC or its affiliates. The RSAC and RSAC CONFERENCE logos and other trademarks are proprietary. All rights reserved.

Steve Wilson

- Exabeam Chief Product Officer
- O’Reilly Author
- OWASP Project Lead

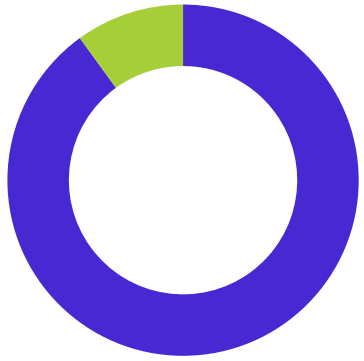
wilsonsd

exabeam

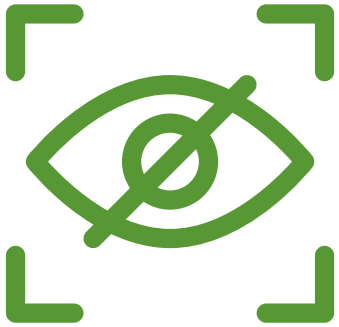


The Machines Are Learning, But Are We?

- Case Studies: A Tale of Two Attacks
  - What happens when your defense fails?
  - A successful defense with AI assistance!
- The Glaring Gaps We See That CISOs Miss
- AI/ML Goes on Offense
- Learning to Use the Latest AI



90% of respondents believe they have good/excellent threat detection ability.



66% On average organizations reported they can only see 66% of their IT environment

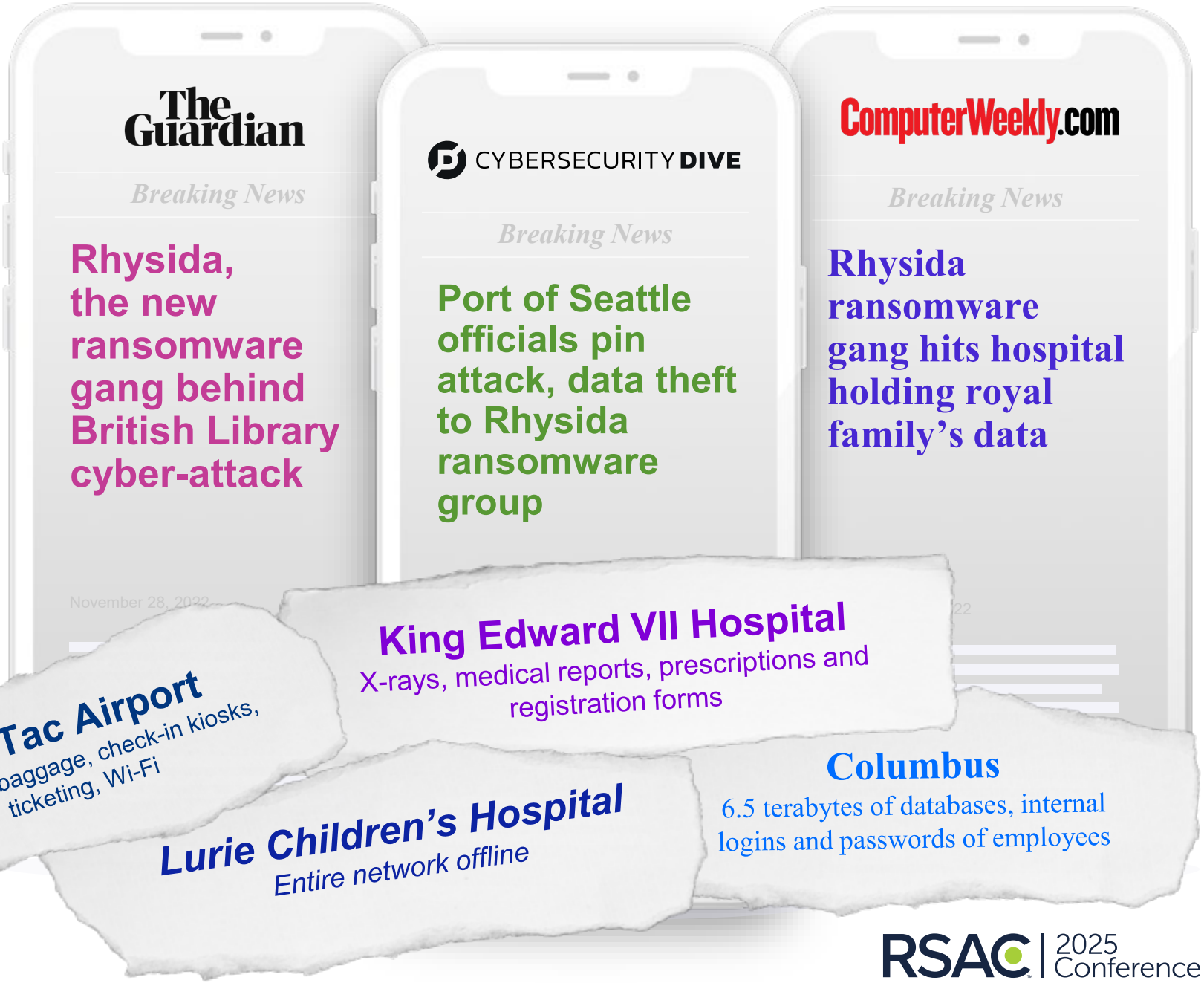
Source: The State of Threat Detection, Investigation, and Response Report November 2023

How did we get to 90% confidence with only 66% visibility?

Source: The State of Threat Detection, Investigation, and Response Report November 2023

Hard Lessons from Rhysida

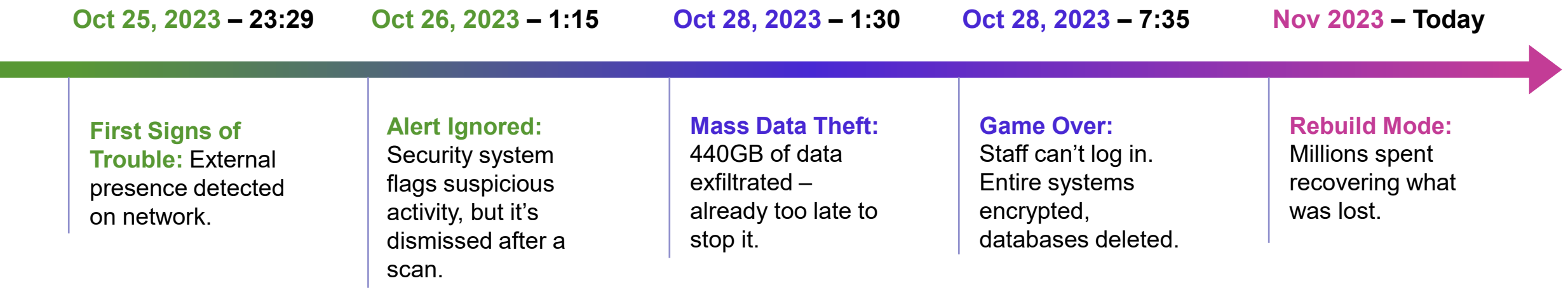
Even well-defended organizations can fail without full visibility into threats.


The Guardian
Breaking News
Rhysida, the new ransomware gang behind British Library cyber-attack
November 28, 2022
CYBERSECURITY DIVE
Breaking News
Port of Seattle officials pin attack, data theft to Rhysida ransomware group
ComputerWeekly.com
Breaking News
Rhysida ransomware gang hits hospital holding royal family's data
SeaTac Airport
Disrupted baggage, check-in kiosks, ticketing, Wi-Fi
King Edward VII Hospital
X-rays, medical reports, prescriptions and registration forms
Lurie Children's Hospital
Entire network offline
Columbus
6.5 terabytes of databases, internal logins and passwords of employees

exabeam

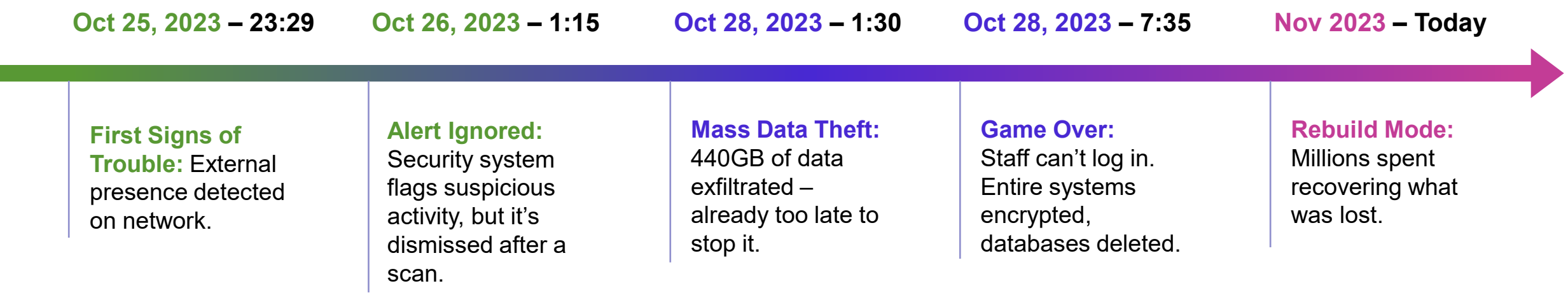
RSAC 2025 Conference

The Horror Story We're All One Click Away From


Oct 25, 2023 – 23:29
First Signs of Trouble: External presence detected on network.
Oct 26, 2023 – 1:15
Alert Ignored: Security system flags suspicious activity, but it's dismissed after a scan.
Oct 28, 2023 – 1:30
Mass Data Theft: 440GB of data exfiltrated – already too late to stop it.
Oct 28, 2023 – 7:35
Game Over: Staff can't log in. Entire systems encrypted, databases deleted.
Nov 2023 – Today
Rebuild Mode: Millions spent recovering what was lost.

Think About What Went Wrong

- ! A suspicious login was detected, but no deeper investigation happened.
- ! The attackers moved laterally across the network, unchecked.
- ! The ransomware attack wasn't stopped until everything was locked down.


Oct 25, 2023 – 23:29
Oct 26, 2023 – 1:15
Oct 28, 2023 – 1:30
Oct 28, 2023 – 7:35
Nov 2023 – Today
First Signs of Trouble: External presence detected on network.
Alert Ignored: Security system flags suspicious activity, but it's dismissed after a scan.
Mass Data Theft: 440GB of data exfiltrated – already too late to stop it.
Game Over: Staff can't log in. Entire systems encrypted, databases deleted.
Rebuild Mode: Millions spent recovering what was lost.

“The Library had modern tools in place, but they were not able to completely monitor or protect our network.”

British Library cyber incident review, 8th March 2024

https://www.bl.uk/home/british-library-cyber-incident-review-8-march-2024.pdf

“Legacy network topology may prevent modern security tools from having full coverage or from being fully effective.

British Library cyber incident review, 8th March 2024

https://www.bl.uk/home/british-library-cyber-incident-review-8-march-2024.pdf

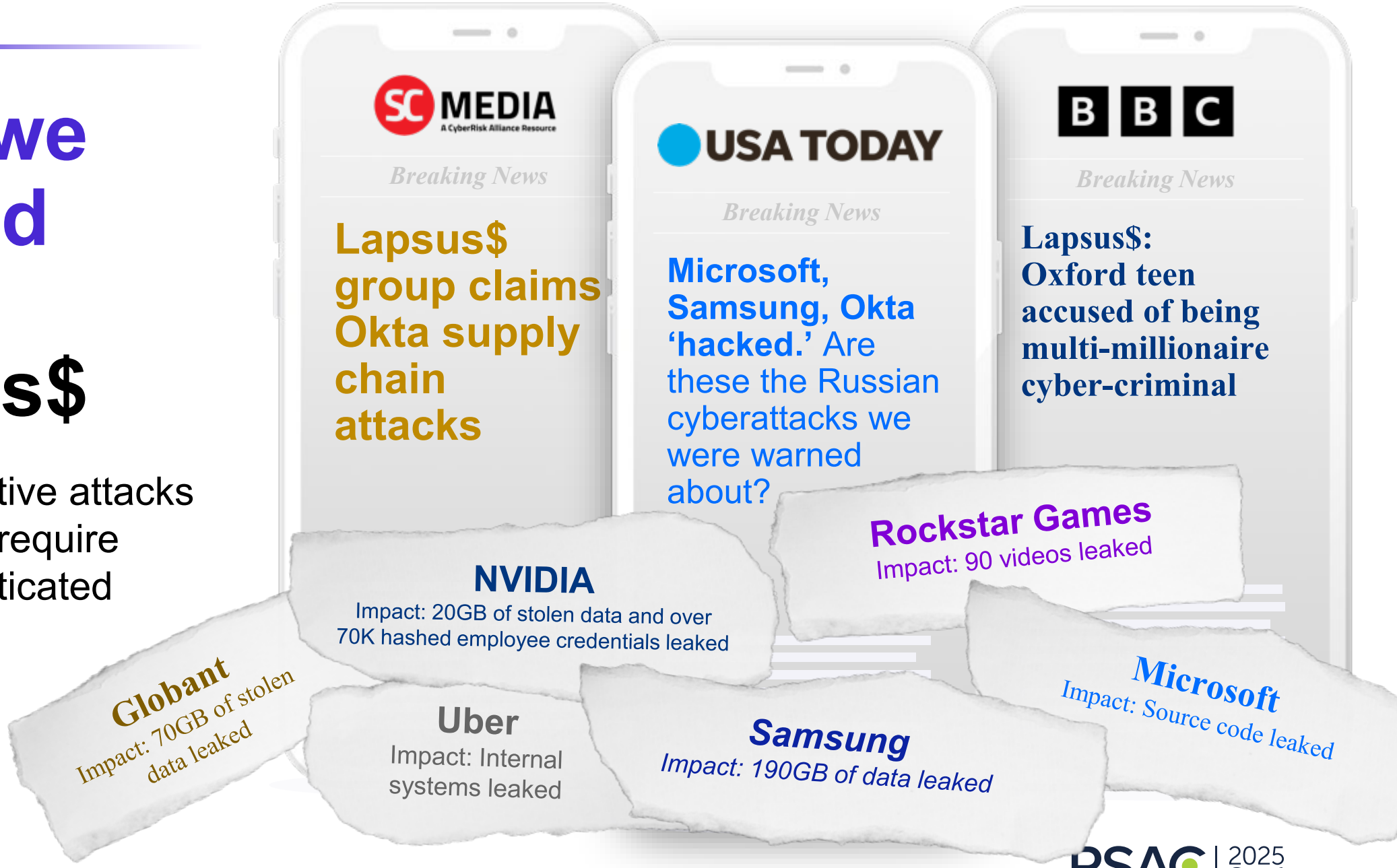
Top of the British Library's recommendations list:

Enhance Network Monitoring Capabilities

The Big Lesson:
You'll Get Few Clues and Limited Time to Respond.
Find the Needle

What we learned from Lapsus$

Highly disruptive attacks don't always require highly sophisticated techniques

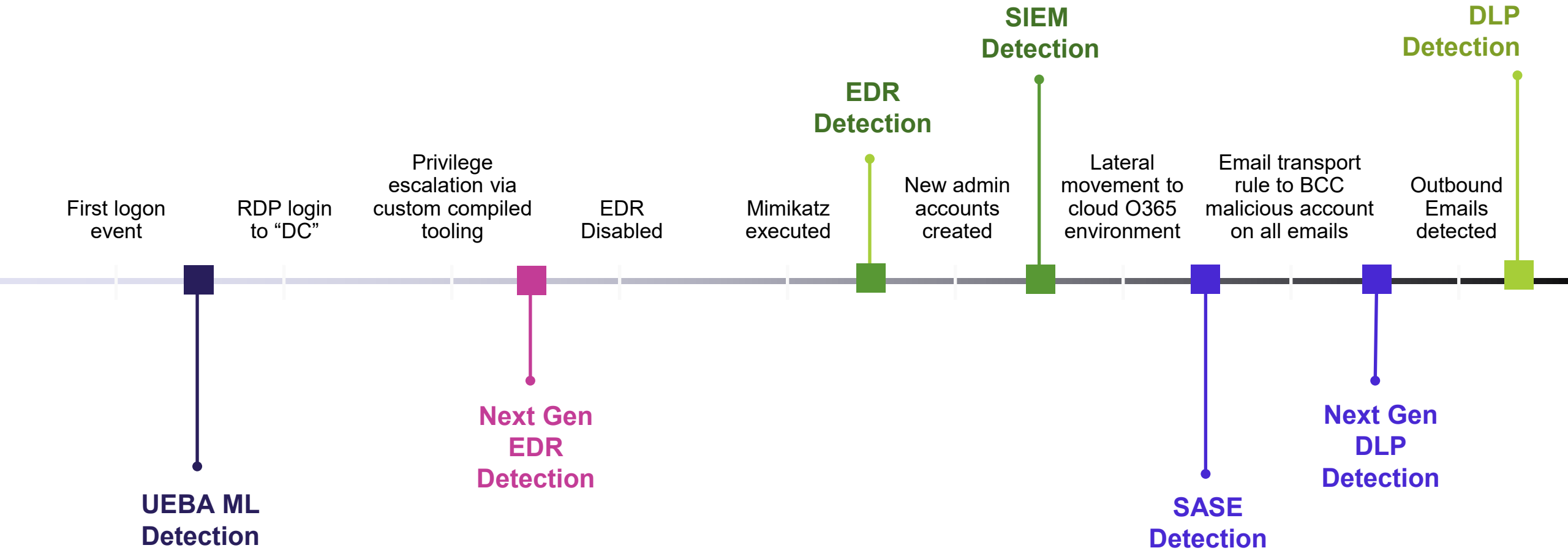


Learning from successful defense?

Let’s Recreate the Attack!

Successfully Defending with Machine Learning

Timeline of a Lapsus$ attack



Compromised Insider: What Are You Not Seeing?

EDR alert to Powershell and Mimikatz on server

Time into attack: **5hr. and 12 min**

May 7, 22:15

Initial access: First login from Singapore for user and org

- First activity from ISP for user
- First OS/browser combination in user agent string

May 7, 23:15

Lateral movement:

# of first access to asset for user: **272**

May 8, 3:20

Lateral movement:

- Attacker pivots off initial host to ServerX
- First RDP activity for user
- First access to ServerX

May 8, 3:27

May 8, 3:27 – 9:56

Lateral movement & privilege escalation:

# of first access to assets by user: 487

# of first communication between assets : 487

# of first credential switches from host: **233**

Important details that were missed

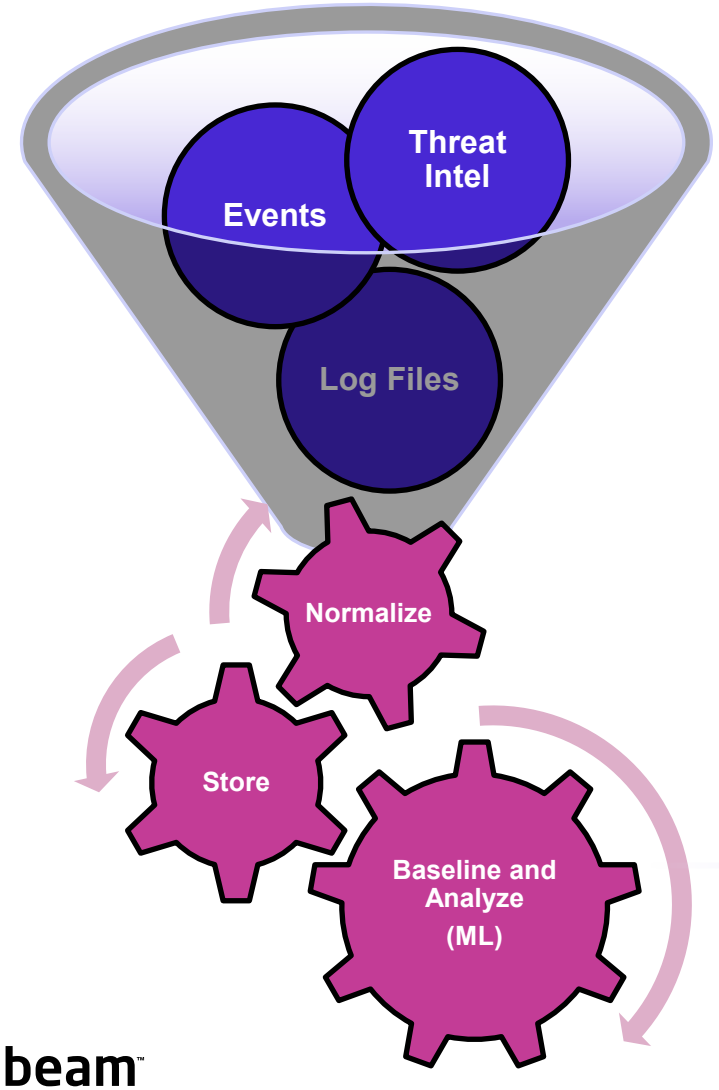
Details of abnormal user behavior

Affected assets: 759

Number of credentials switched: 234

notepad.exe running out of an abnormal directory C:\\PerfLogs

Reducing Petabytes of Data into Kilobytes of Insight with ML



TIMELINE

1
- ✓ Log-on
- ✓ Risk Score: None

2
- ✓ Unusual Access
- ✓ Risk Score: Low

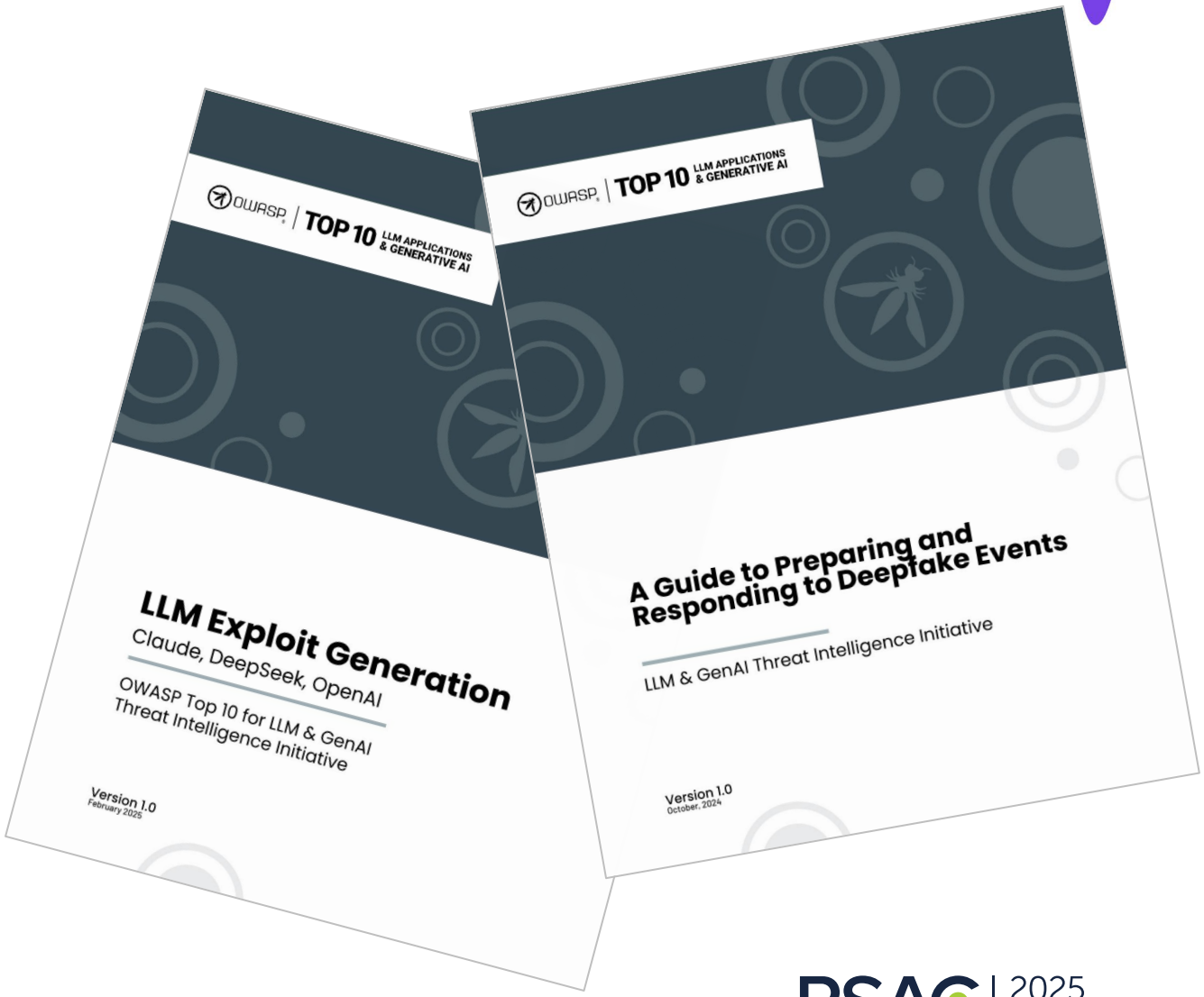
3
- ✓ Possible Exfiltration
- ✓ Risk Score: High

ML-Powered Behavior Analytics

Found and Stopped Attack Hours Earlier!

The Bad Guys Have Frontier AI

- Hackers are using OpenAI & Google
- Now have unrestricted models from China
- DeepFakes are real
- Exploit generation is picking up
- Time to learn and evolve!


OWASP | TOP 10 LLM APPLICATIONS & GENERATIVE AI
LLM Exploit Generation
Claude, DeepSeek, OpenAI
OWASP Top 10 for LLM & GenAI Threat Intelligence Initiative
Version 1.0
February 2025
OWASP | TOP 10 LLM APPLICATIONS & GENERATIVE AI
A Guide to Preparing and Responding to Deepfake Events
LLM & GenAI Threat Intelligence Initiative
Version 1.0
October, 2024

We are too slow because we are still over-reliant on humans for TDIR

46% of respondents have automated more than half of their TDIR workflow

Source: The State of Threat Detection, Investigation, and Response Report November 2023

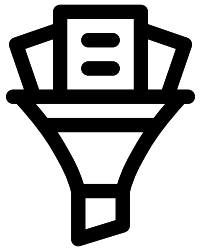
The Challenges Speak Volumes

Automate this	Time-consuming investigation processes	42%
Automate this	Insufficient threat intelligence	41%
Automate this	Tooling is too complex	40%
Already automation	Lack of automation across TDIR workflow	39%
So you have *TIME* for this	Lack of knowledge on how to respond to incidents	38%
Automate this	Limited visibility into IT environment	35%
Automate this	Lack of skilled personnel	32%
Automate this	Reporting is not useful/has to be manually done for stakeholders	32%

Optimizing AI for Effective Security Operations

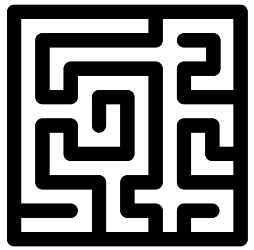
- ✓ Define Clear Use Cases and Objectives
- ✓ Build Trust with Explainable AI
- ✓ Automate Investigations and Agentic AI
- ✓ Iterate & Improve

The Roadblocks to Smarter Security

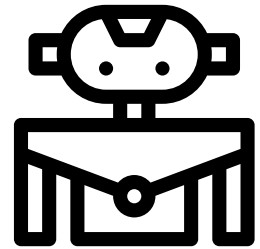


Data Silos

Human Over-Reliance



Legacy Infrastructure



Lack of Confidence in AI

Analysts Have Too Many Questions to Answer Manually

1 event = 46 questions

- Who owns this asset?
- What is installed on it?
- Is it a business role?
- Any recent changes?
- What host does this IP map to and for how long?
- What user does this IP map to and for how long?

CMDB/Directory Services/Operating System/DHCP

- Is this port opened?
- Is it authorized?
- What is it used for?
- Is this normal traffic behavior?
- Have these hosts communicated with each other using this pattern before?
- What phase in the cyber kill chain (recon) is it?

Traffic Analysis, Ports/Protocols

- What does this alert mean?
- How does it work/what makes it fire?

Alert Details

- Who is this user?
- What is their status?
- What is their role?
- How does their activity compare to their peers/org?
- What privileges do they have?
- What groups do they belong to?
- What is their contact info.?
- Has this user connected to these hosts before?

Directory Services/HRMS

May 2 2019 11:49:00 host1 10.78.121.42:350 10.28.161.16:203 up.badsite.local/upload.jar Large outbound traffic volume user=bsalazar winscp.exe

Historical and Current Info.

- How long has this activity occurred?
- What else is happening?
- Is this an approved time period?

DIG

- Where is this going to?

Domain Tools

- What are these hosts names?

Threat Intel

- Is this a known bad actor?
- Have we seen this address accessed by any user/peer group/ the org before?
- What phase in the cyber kill chain (recon) is it?

Machine Learning/Analytics

- Have any of our featured classification algorithms identified this as malicious?
- What is the entropy score for this URL, for the domain?
- Have we seen any user/peer group/the org visit this site before?
- What phase in the cyber kill chain (delivery/payload, C&C) is it?

CVE/Open Source/ Commercial/Internal Intel

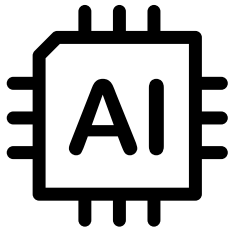
- How new is this domain?
- Is this a known indicator of compromise?
- What is the risk rating /reputation of the domain?
- Is this domain known to serve up malicious content?
- Is this URL being reported as malicious?
- Is this an exploit call or known common exfiltrate call?
- What phase in the cyber kill chain (exploit) is it?

Processes

- Is this an authorized process?
- What is it used for?
- Have we seen this before from the user/peer/in the org?
- What is the file hash?
- What phase in the cyber kill chain (install, action/objective) is it?

Building Agents with GenAI

Embedded Security Knowledge



LLM

Natural Language

- Overview
- Possible Threats
- Next Steps
- …
- Follow-up
- Discussion

Detailed Threat Context

AI as an Analyst Force Multiplier

Traditional Security Teams	AI-Enabled Security Teams
🚨 Too many alerts to manually investigate	🤖 AI automatically investigates anomalies
🕵️‍♂️ Analysts spend hours piecing together data	📊 AI generates a full attack timeline instantly
⏳ Response times are delayed due to manual processes	🔥 AI takes immediate action (isolates assets, revokes access, updates security policies)
❌ Attackers move freely across the network	🚪 Network segmentation is enforced automatically
😵 Security teams are overwhelmed and reactive	🚀 Teams focus on strategy, not firefighting

Two Force Multiplier Perspectives

“The agent is halving the time it takes me to do an investigation. Typical investigations would take me 30-45 minutes, with the agent I’m closing them down in about 15 minutes”

“It’s like having an enthusiastic undergrad on hand to provide them a detailed technical deep dive, so you don’t have to do it yourself”

Call to Action: Are We Learning?

The machines are learning at an unprecedented rate. Are we?

Our research shows:

- Confidence without capability is a vulnerability.
- Organizations that succeed will be those that continuously adapt.
- Leverage AI not just to respond to threats, but to predict and prevent them.

This is the new cybersecurity paradigm.

Cybersecurity leaders must:

- Embrace AI-driven defenses or risk falling behind.
- Invest in behavioral analytics, AI agents, and automated threat detection.
- Prepare for adversaries who are already leveraging AI in attacks.

Apply: What Should You Do Next?

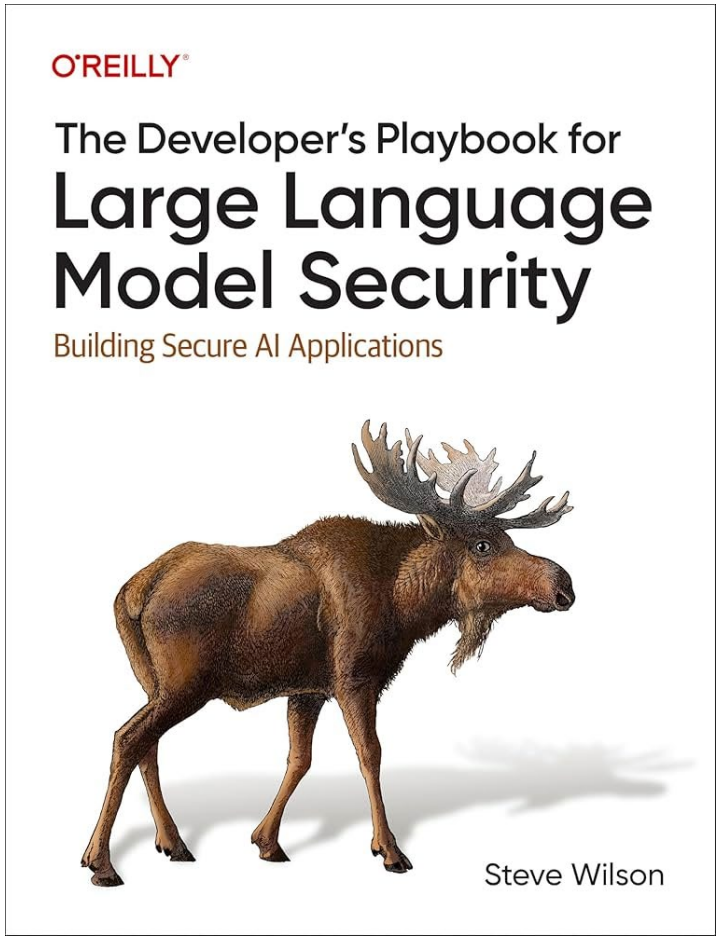
Immediate Actions (Now)

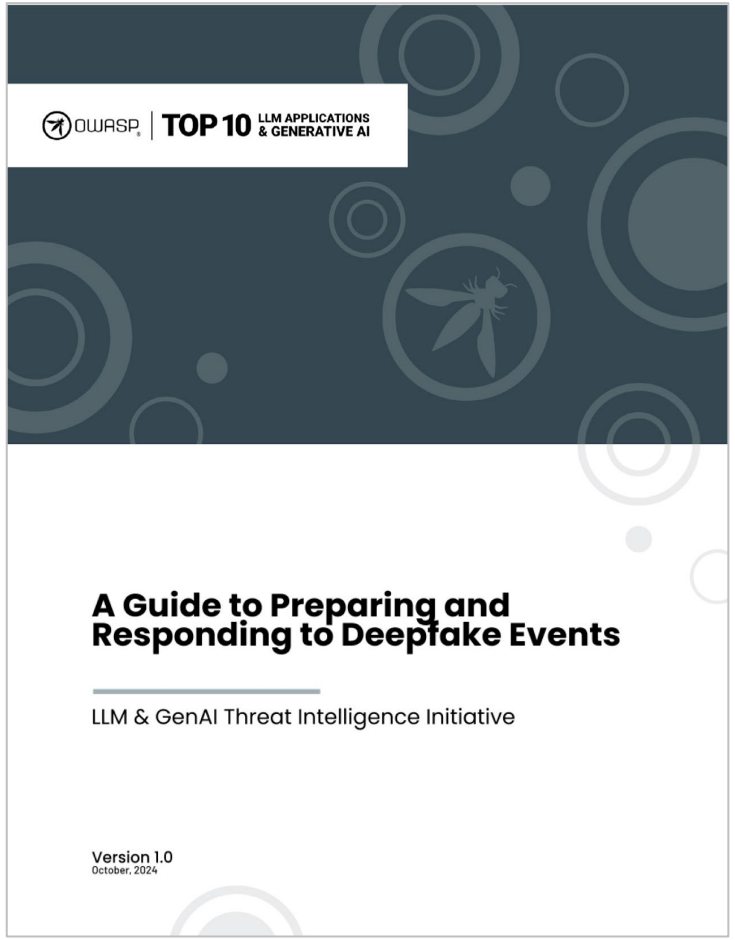
1. Assess your AI-driven security readiness
2. Engage with your security team on AI risks

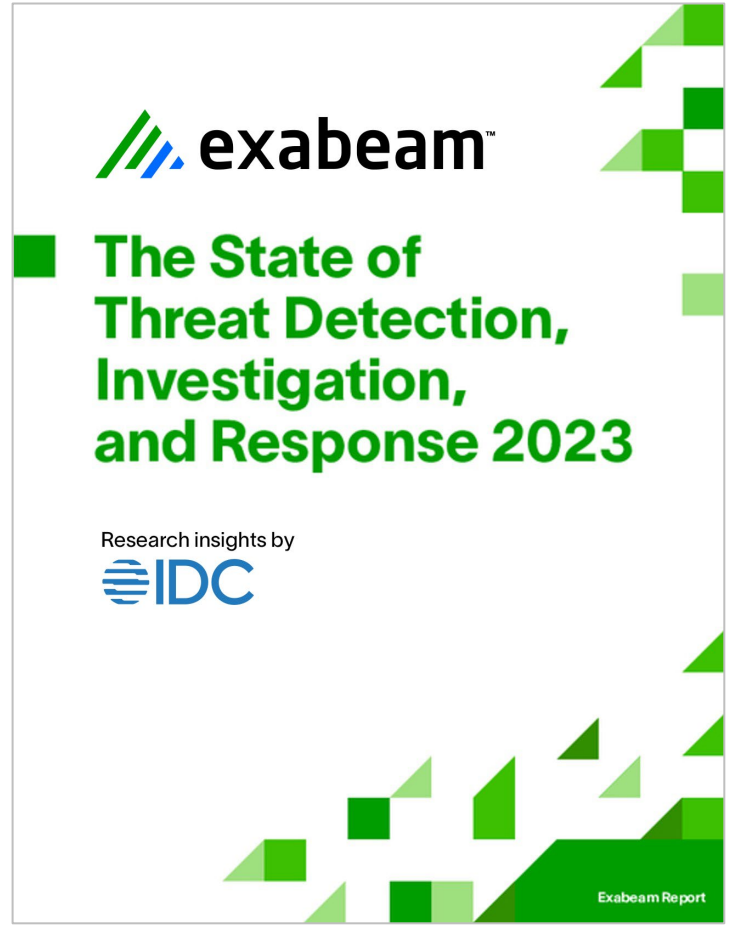
Actions to Implement Within 3 Months

1. Conduct an AI-Driven Threat Simulation
2. Improve Visibility & Data Readiness for AI Security
3. Automate Low-Level Security Investigations
4. Prepare for AI-Augmented Attacks
5. Collaborate & Share Intelligence

Further Reading




OWASP | TOP 10 LLM APPLICATIONS & GENERATIVE AI
A Guide to Preparing and Responding to Deepfake Events
LLM & GenAI Threat Intelligence Initiative
Version 1.0
October, 2024



The machines are learning,
so should we!
exabeam
RSAC 2025 Conference

RSAC 2025 Conference

Many Voices.
One Community.

Thank You

For more information:

Please reach out on LinkedIn

wilsonsd